第 29 卷 第 4 期　　昆 虫 学 报　　Vol. 29, No. 4
1986 年 11 月　　ACTA ENTOMOLOGICA SINICA　　Nov., 1986

菜蚜种群抽样理论的蒙特卡洛试验研究

沈佐锐　管致和
(北京农业大学植保系)

J. Deltour　P. Dagnelie
(比利时 Gembloux 农学院)

摘要　北京地区秋白菜上蚜虫种群在其一定发展阶段的空间格局是可以用 Pearson III 型分布拟合的。用蒙特卡洛试验研究这种分布的三个参数——平均数 Ex，变异系数 Cv 和偏态系数 Cs 的估计误差同样本容量的关系是有效的途径。本文介绍了该试验的设计思想和实施步骤，并得出初步结论：以一株菜为一个样本单位计数白菜上的蚜虫头数，则在允许误差不大于 5% 的情况下，样本容量为 50 时，用矩法便可足够准确地估计 Ex 和 Cv 值了；但对于估计 Cs 值，则样本容量应为 500 左右。

关键词　蚜虫种群　空间分布　皮尔逊曲线　抽样　蒙特卡洛法

一、做为菜蚜种群空间格局模型的 Pearson III 型分布

蚜虫做为典型的 r-策略昆虫，其种群在数量动态和空间格局上的特点是：①各种数量特征不稳定，变化较快；②在发生后不久便可达到相当大的种群数量；③样本中个体数变量的离散程度较高。这些特点是在兼性多型、世代重叠和有翅蚜迁飞等生物学特点的基础上形成的。因此，用传统的 Poisson 分布、负二项分布、Neyman A 型分布、Poisson-二项分布等离散型概率分布来拟合菜蚜种群空间抽样数据显然是行不通的。根据Kendall 和 Stuart（1977）提出的拟合复杂观测频率分布的三种主要途径，沈佐锐等（1985）初步证明，利用 Pearson 曲线系统可以描述北京地区秋白菜上桃蚜 *Myzus persicae* (Sulzer)和萝卜蚜 *Lipaphis erysimi* (Kaltenbach) 混生种群的空间分布型动态，而其中 III 型曲线则可描述菜蚜种群在其一定发展时期所呈现的空间分布状态。这里系以一株菜为一个样本单位计数每株菜上蚜虫头数。

Pearson III 型分布的概率密度函数为

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}(x-b)^{\alpha-1}e^{-\beta(x-b)} \quad (b \leqslant x < \infty)$$

式中 α、β、b 为原始参数；$\alpha, \beta > 0$；当 $b = 0$，成为 Γ-分布。

Pearson III 型分布的曲线形状又可由三个数字特征决定：数学期望 Ex，变异系数

本文于 1985 年 4 月收到。
林昌善教授、陈维博同志和徐汝梅同志为本文提出了宝贵的意见，丁岩钦副研究员一再鼓励本文的写作和发表，在此一并致谢。

Cv 和偏态系数 Cs；故三者亦称形状参数。原始参数和形状参数的关系是

$$\alpha = \frac{4}{C_s^2}, \quad \beta = \frac{2}{Ex \cdot Cv \cdot Cs} \quad 和 \quad b = Ex \cdot \left(1 - \frac{2Cv}{Cs}\right)$$

当从 Pearson III 型分布总体中抽样时，我们以矩法计算 Ex，Cv 和 Cs 的估计量：

$$\hat{E}x = \frac{1}{N}\sum x_i = \bar{x}$$

$$\hat{C}v = \sqrt{\frac{\sum(K_i - 1)^2}{N-1}}$$

$$\hat{C}s = \frac{\sum(K_i - 1)^3}{(N-3) \cdot Cv}$$

式中$\sum$表示 $\sum_{i=1}^{N}$，$K_i = \frac{x_i}{\bar{x}}$。

二、蒙特卡洛试验的设计思想

有害生物单种种群的研究存在着两个方向，一个是种群数量动态，一个是种群空间格局。就前者而言，人们关心的主要是对 Ex 的估计，种群数量或密度直接关系到有害生物所造成的经济损失和人类的防治策略；就后者而言，为了研究一个样本是否可用 Pearson III 型分布拟合，还要同时考虑对 Cv 和 Cs 的估计，用解析的方法来解决这个问题是相当困难的。然而这恰是蒙特卡洛方法能够显示其优越性的场合，即使对 Ex 的估计也许存在着数据变换然后用公式求算的方法，但那样做比较适合于只有一个或两个参数的分布（丁岩钦，1980；Ruesink, 1980），而对 Pearson III 型分布来说，仍不妨用蒙特卡洛方法讨论。

蒙特卡洛试验的原理和步骤

1. 首先确定试验所要求的真值，即 Pearson III 型分布的总体参数 Ex_0，Cv_0 和 Cs_0。然后在计算机上产生 Pearson III 型分布变量，这相当于我们在一个 Pearson III 型分布总体中抽样。每产生一个变量即抽得一个样本单位，每进行一次试验即得到一个样本。在一次试验中所产生的变量总数则相当于样本容量。

2. 每次试验结束后，计算每个总体参数的样本估计量。在参数估计的三种常用方法——矩法、最大或然法和适线法中，我们采用矩法。

3. 重复试验　所重复的试验次数应使样本估计量的平均值达到试验要求的精度。

4. 设计不同的样本容量水平　对每一水平的样本容量进行了上述三步后，逐个计算三个参数在各次试验中的估计量之平均值，并与真值比较，计算相对误差。

5. 以相对误差评价各参数估计所要求的理论抽样数。

为了在计算机上产生 Pearson III 型分布变量，我们参考了华东水利学院（1980）介绍的框图；这是 1973 年 Whittaker 提出的舍选抽样方法，具有较高的抽样效率。

整个试验的程序用 TRS-80 微型机的 BASIC Level II 语言写成，该程序所需的均匀分布变量由该机的内部函数 RND 产生。

为更真实地模拟菜蚜种群在田间的分布，在上述第一步中增加两个措施：a. 所输入

的真值是以田间实测数据为基础的，即通过适线法用 Pearson III 型分布拟合实测数据，如此得到的 Ex，Cv 和 Cs 估计值作为蒙特卡洛试验的输入真值；b. 由计算机产生的 Pearson III 型分布变量本是正实数，我们作了整数化处理。

三、蒙特卡洛法本身的精度问题

设 η 为被模拟的 Pearson III 型分布总体参数的样本估计量，这是随机变量，$E\eta$，$\sigma\eta$ 为其数学期望与方差，L 是抽取的样本数量。我们用 η 的样本平均数

$$m\eta = \frac{1}{\mathrm{L}} \sum_{i=1}^{\mathrm{L}} \eta_i \tag{1}$$

来估计 $E\eta$，当 $\mathrm{L}\to\infty$ 时，不论 η 的分布如何，$(m\eta - E\eta)$ 趋于正态分布 $N(0, E\eta/\sqrt{\mathrm{L}})$。于是有

$$P\{|m\eta - E\eta| < \varepsilon\} = \int_{-\frac{\varepsilon\sqrt{\mathrm{L}}}{\sigma\eta}}^{\frac{\varepsilon\sqrt{\mathrm{L}}}{\sigma\eta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt。 \tag{2}$$

在 95.5% 置信水平上，

$$\frac{\varepsilon\sqrt{\mathrm{L}}}{\sigma\eta} = 2。 \tag{3}$$

假定允许误差为 $e\%$，即要求

$$\frac{\varepsilon}{E\eta} \leqslant e\%, \tag{4}$$

则由式 (3) 有

$$L \geqslant \left(\frac{200}{e}\right)^2 \cdot Cv\eta^2, \tag{5}$$

式中

$$Cv\eta = \frac{\sigma\eta}{E\eta} \tag{6}$$

为 η 的变异系数(从树铮等，1980)。

在开始蒙特卡洛试验时，我们并不知道 $Cv\eta$ 的值，可通过下面两个阶段性试验解决：

(1) 预备阶段试验

设 L = 50，并设样本容量有 EN = 50，100，200，300，400 五个水平。给定总体参数后，即进行 Pearson III 型分布抽样。得到的每个样本用矩法估计 Ex、Cv 和 Cs。这样，每个样本容量水平的 50 个样本都可给出这三个估计量的平均数和变异系数，结果列于表 1。

根据预备阶段试验的结果，可粗略估计 L 值，即先以 Vx、Vv 和 Vs 为 $Cv\eta$ 分别代入式 (5)，并令 $e = 5$ 求出相应的 L 值后再适当放大，列入表 2。

(2) 检验阶段试验

表 1　预备阶段试验的结果*

样本	总体参数	EN	Ex	Vx	Cv	Vv	Cs	Vs
9.14	$Ex_0=19.6692$ $Cv_0=0.8008$ $Cs_0=1.9263$	50	19.4926	0.0974	0.7785	0.1446	1.6353	0.4153
		100	19.3942	0.0738	0.7986	0.0903	1.8331	0.2604
		200	19.7592	0.0577	0.8077	0.0781	1.8724	0.2655
		300	19.5237	0.0519	0.8065	0.0565	1.8821	0.2204
		400	19.6273	0.0332	0.7992	0.0527	1.9055	0.2041
10.2	$Ex_0=596.221$ $Cv_0=0.7313$ $Cs_0=1.8181$	50	593.424	0.0955	0.7267	0.1581	1.7146	0.4127
		100	601.352	0.0854	0.7278	0.1014	1.7825	0.3717
		200	597.416	0.0474	0.7299	0.0675	1.7120	0.2384
		300	598.542	0.0335	0.7420	0.0574	1.8628	0.2024
		400	596.127	0.0311	0.7317	0.0424	1.7333	0.1368

* 表头中 Ex、Cv、Cs 分别为由 L＝50 个样本的 Ex、Cv 和 Cs 计算得到的平均数（$m\eta$），Vx、Vv 和 Vs 为相应的 $Cv\eta$ 的计算值。

表 2　5%允许误差下的 L 值

样本	EN	L_x		L_v		L_s	
		计算值	估定值	计算值	估定值	计算值	估定值
9.14	50	15.2	50	33.45	70	275.95	300
	100	8.71	40	13.05	50	108.49	250
	200	5.33	40	9.759	50	112.78	150
	300	4.31	30	5.108	40	77.722	100
	400	1.76	30	4.444	40	66.651	80
10.2	50	14.6	50	39.99	80	272.51	300
	100	11.7	50	16.45	70	221.05	250
	200	3.59	30	7.290	50	90.935	150
	300	1.80	30	5.272	40	65.545	100
	400	1.55	30	2.876	40	29.943	80

表 3　检验阶段试验结果之一例（计算机打印输出）

SAMPLE 9.14

(1) POPULATION PARAMETERS OF P-III DISTRIBUTION:

$Ex=19.6692$ $Cv=0.8008$ $Cs=1.9725$

(2) ORIGINAL AND SECONDARY PARAMETERS:

AF ＝ 1.0281 BI ＝ 3.6985 BT ＝ 0.0644 NP ＝ 1 OP ＝ 0.0281

EN	L	Ex	Vx	Cv	Vv	Cs	Vs	e*
50	300	19.6447	0.1079	0.7781	0.1330	1.6291	0.3653	4.2
100	250	19.6878	0.0827	0.7938	0.1013	1.7451	0.2948	3.7
200	150	19.6416	0.0524	0.7976	0.0729	1.8779	0.2318	3.8
300	100	19.5650	0.0409	0.8026	0.0622	1.9390	0.2400	4.8
400	80	19.6204	0.0419	0.7972	0.0482	1.8970	0.1966	4.4
500	70	19.6661	0.0392	0.7997	0.0415	1.9310	0.1648	3.9

* 对于 Vx、Vv，其计算出的 e 值显然小于 Vs 计算的 e 值，故未列入表内。

以表 2 给出的各 L 值代替预备阶段试验中的 L = 50，重复预备阶段试验的步骤，然后以式(5)计算 e_o 试验至少重复两次。表 3 为其中一次计算机输出结果。

然后我们就检验阶段的两次试验结果检验所取的 L 值是否合适。如果在某个水平上某参数估计的相对误差大于 5%，则这个水平的 L 值应再放大一些。结果表明，我们这里选取的 L 值都能满足蒙特卡洛试验本身的精度要求。

四、结果

各样本容量水平的参数估计相对误差可用下式计算:

$$e_r = \max\left\{\frac{|E\eta - (m\eta \pm 1.96Sm/\sqrt{L})|}{E\eta}\right\} \times 100\%$$

式中，1.96 是 $\alpha = 0.05$ 时正态分布的双侧分位数，$m\eta$ 和 Sm 在计算时分别以检验阶段中两次试验结果合并后的 Ex、Cv 和 Cs 及 Vx、Vv 和 Vs 代替，即

$$m\eta = \frac{m\eta_1 + m\eta_2}{2} \quad 和 \quad Sm = \sqrt{\frac{Sm_1^2 + Sm_2^2}{2}},$$

其中 $Sm_i = m\eta_i \cdot Cv\eta_i$。此结果列于表 4。

表 4　参数估计准确度与样本容量的关系（95%置信水平）

EN	e_r 值(%)					
	对于 Ex		对于 Cv		对于 Cs	
	9.14	10.2	9.14	10.2	9.14	10.2
50	1.42	1.43	4.17	2.48	20.41	16.00
100	1.03	1.20	2.09	2.09	13.70	11.24
200	1.01	1.87	1.95	2.26	9.18	9.15
300	0.92	2.02	1.74	2.27	7.50	7.62
400	0.93	1.91	1.40	2.48	7.08	9.83
500	1.14	1.41	0.97	0.51	5.07	6.38

表 4 指出，在允许误差不大于 5% 的情况下，样本容量为 50 时，用矩法便可足够准确地估计 Ex 和 Cv 值了；但对于估计 Cs 值，则样本容量应为 500 左右。

五、问题与讨论

从表 4 可见:

1. 仅就估计 Ex 而论，样本容量还可更小些。

2. Pearson III 型分布总体参数的不同，对参数估计的准确度亦有影响。

以上两个问题应在进一步的试验中综合研究。

3. 在样本容量小于 500 的情况下，用矩法计算的 Cs 值相对误差大于 5%，故应改用更为精确的方法，如适线法，但这至少要拥有一台微型机(沈佐锐等，1985)。

4. 丛树铮等 (1980) 在用蒙特卡洛试验研究了 Pearson III 型分布参数的估计方法后指出，矩法、极大或然法和适线法对 Ex、Cv 和 Cs 的估计精度不同。所以，当我们为了研究用适线法估计参数的问题而进行蒙特卡洛试验时，上面介绍中用到矩法的地方都

应改为适线法。

5. 试验速度评价以表 3 为例，进行完表中全部计算（从 EN = 50 至 EN = 500），其过程需要连续进行将近 40 小时，这是就 TRS-80 机而言。由于占机时间较长，就限制了蒙特卡洛试验中一些最优化课题的研究。在这种情况下，最好应用大、中型计算机。

参　考　文　献

丁岩钦　1980　昆虫种群数学生态学原理及应用。科学出版社。
丛树铮等　1980　水文频率计算中参数估计方法的统计试验研究。水利学报 3: 1—15。
华东水利学院主编　1980　水文学的概率统计基础。水利出版社。369–370 页。
沈佐锐、管致和、P. Dagnelie 和 J. Deltour 1985 用 Pearson III 型曲线拟合菜蚜种群空间分布数据的初步研究。生态学报 5(4): 364–72。
Kendall, M. and A. Stuart 1977 The Advanced Theory of Statistics, Vol. 1: Distribution Theory. 4th. ed. London. Sec. 6.1.
Ruesink, W. G. 1980 Introduction to Sampling Theory in *Sampling Methods in Soybean Entomology* (eds. Kogan, M. and D. C. Herzog). Springer-Verlag, New York. pp. 61–78.

STUDY ON SAMPLING FROM SIMULATED APHID POPULATION BY MONTE CARLO TEST

SHEN ZUO-RUI　GUAN ZHI-HE
(Department of Plant Protection Beijing Agricultural University, Beijing)

J. DELTOUR　P. DAGNELIE
(Faculté des Sciences Agronomiques de l'Etat, 5800 Gembloux, Belgique)

For mixed populations of *Myzus persicae* (Sulzer) and *Lipaphis erysimi* (Kaltenbach) on Chinese cabbage in autumn in the area of Beijing, some of their spatial distributions can be simulated with the type III of Pearson curve system (P-III distribution), whose variables can be generated in terms of the Monte Carlo method on TRS-80 microcomputer. The P-III distribution is definited with three shape parameters: Ex, the mean; Cv, the coefficient of variation and Cs, the coefficient of skewness. The present paper deals with relation between acouracy of the parameter estimates and sample size from the P-III distribution population generated by the method in Whittaker (1973).

Let η denote the estimates of Ex, Cv and Cs; $E\eta$, $a\eta$ respectively mathematical expects and variances of the estimates; L the number of samples to enter into statistical calculation; and e the permissible relative error, thus

$$L \geqslant \left(\frac{200}{e}\right)^2 Cv\eta^2,$$

where

$$Cv\eta = \frac{\sigma\eta}{E\eta}.$$

In order to ensure the precision of Monte Carlo test, the L value should be determinated. So firstly, an exploratory test is made:

Let L=50 and sample size has five levels, that is EN=50, 100, 200, 300, 400. At the end of the test, $Cv\eta$-values are estimated by the moment method and then L values are gained when $e=5\%$.

Secondly, the L values are verified.

Finally, relative errors in the parameter estimation, corresponding to the levels of sample size, are calculated by

$$e_r = \max\left\{\frac{|E\eta - (m\eta \pm 1.96Sm/\sqrt{L})|}{E\eta}\right\} \times 100\%$$

The result shows that in the case of e=5%, the moment method can quite exactly estimate Ex and Cv even when EN=50; but for Cs, the method requires EN=500 at least.

Key words Aphid population——Sampling——Monte Carlo——Pearson curves——Spatial distribution